

Splicing in *Caenorhabditis elegans* Does Not Require an AG at the 3' Splice Acceptor Site

RAFFI V. AROIAN,^{1†} ADAM D. LEVY,² MAKOTO KOGA,³ YASUMI OHSHIMA,³
JAMES M. KRAMER,² AND PAUL W. STERNBERG^{1*}

Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, Pasadena, California 91125¹; Department of Cell, Molecular, and Structural Biology, Northwestern University, Chicago, Illinois 60611²; and Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan³

Received 5 May 1992/Returned for modification 5 August 1992/Accepted 28 October 1992

The dinucleotide AG, found at the 3' end of virtually all eukaryotic pre-mRNA introns, is thought to be essential for splicing. Reduction-of-function mutations in two *Caenorhabditis elegans* genes, the receptor tyrosine kinase gene *let-23* and the collagen gene *dpy-10*, both alter the AG at the end of a short (ca. 50-nucleotide) intron to AA. The in vivo effects of these mutations were studied by sequencing polymerase chain reaction-amplified reverse-transcribed RNA isolated from the two mutants. As expected, we find transcripts that splice to a cryptic AG, skip an exon, and retain an unspliced intron. However, we also find significant levels of splicing at the mutated 3' splice site (AA) and at nearby non-AG dinucleotides. Our results indicate that for short *C. elegans* introns an AG is not required for splicing at either the correct 3' splice site or incorrect sites. Analysis of a splice site mutant involving a longer, 316-nucleotide *C. elegans* intron indicates that an AG is also not required there for splicing. We hypothesize that elements besides the invariant AG, e.g., an A-U-rich region, a UUUC motif, and/or a potential branch point sequence, are directing the selection of the 3' splice site and that in wild-type genes these elements cooperate so that proper splicing occurs.

The dinucleotide AG is an essential feature of the 3' splice acceptor site of eukaryotic pre-mRNA introns. A comparison of approximately 3,700 pre-mRNA introns from vertebrates, invertebrates, plants, yeasts, and viruses shows that the AG at the 3' end of introns is virtually invariant: all contain the ultimate nucleotide G, and only four exceptions (involving alternative splice sites) to the penultimate nucleotide A were found (27, 50). All the introns of the nematode *Caenorhabditis elegans* analyzed to date also end with the dinucleotide AG (15).

Consistent with its striking conservation, the invariant AG is functionally critical for proper splicing. Alteration of the invariant AG to either AA, AU, CG, GG, or UG inhibits splicing in vitro (1, 35, 45). The importance of the dinucleotide AG is also supported by in vivo studies. Introns harboring point mutations in the AG dinucleotide are abnormally spliced in yeasts (1), *Neurospora crassa* (34), and mammalian cells (25, 38). Conversely, mutations that create an AG (e.g., AC mutated to AG) sometimes result in the selection of the newly created AG as a new 3' splice site (36, 57).

The necessity of having an AG at the 3' end of introns also has implications for numerous human disorders. Diseases associated with point mutation of either the A or the G at the 3' end of an intron include analbuminemia (47), β -thalassemia (4), citrullinemia (30, 52), cystic fibrosis (21), debrisoquine polymorphism (22), hemophilia (9), familial type III hyperlipoproteinemia (10), lipoprotein lipase deficiency (23), ornithine transcarbamylase deficiency (8), osteogenesis imperfecta (58), Tay-Sachs disease (39), and xeroderma pigmentosum (49).

The in vivo consequences of altering the AG at a 3' splice

site vary and generally fall into three classes: cryptic splicing, exon skipping, and a failure to splice. When the AG at the end of an intron is mutated or deleted, cryptic splicing can occur at an AG that is not normally used as a 3' splice site. This cryptic AG is usually, but not always, nearby and can be located downstream (1, 30, 34, 49, 52) or upstream of the original 3' splice site (4, 10, 25). On the other hand, mutation or removal of the AG at the 3' end of an intron can sometimes result in exon skipping: the 5' end of the mutated intron is used as the donor splice site, and the 3' end of the next intron downstream is used as the acceptor splice site (38, 58). As a result, the mutant intron, the next exon, and the next intron are all spliced out together. Lastly, an intron harboring a mutated or deleted AG sometimes remains unspliced (10, 52).

Here we report on the splicing defects associated with mutation of the invariant AG to an AA in two short *C. elegans* introns: the 53-nucleotide (nt) intron 16 of the *let-23* gene and the 48-nt intron 2 of the *dpy-10* gene. Both the number and the nature of the resulting splicing products are unprecedented. We have also analyzed a double AG to AA mutation in a longer 316-nt intron of *let-23*. Our results indicate that splicing in *C. elegans* does not require an AG at the 3' splice site, suggesting that elements other than an AG can direct the selection and usage of 3' splice sites. We discuss the implications of these findings to splicing in *C. elegans* and other organisms.

MATERIALS AND METHODS

We describe the materials and methods used for the *let-23* and *dpy-10* investigations separately since the studies were mostly carried out independently.

Localization of mutations. (i) *let-23(n1045)* and *let-23(sy97sy122)*. The *n1045* mutation was localized and sequenced as described for the allele *let-23(sy5)* (2). The *n1045* intron 16 mutation discussed here is most likely the only

* Corresponding author.

† Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448.

mutation associated with *n1045*. The entire *let-23(n1045)* coding sequence and all exon/intron boundaries were examined multiple times for hydroxylamine polymorphisms, and only the intron 16 polymorphism was found. Furthermore, both genomic and cDNA sequence have been generated for more than 100 bases of either side of the *n1045* mutation. No other mutation has been found. The mutation *sy122* was isolated as an intragenic revertant of the *let-23* allele *sy97* (27b). The specific alteration associated with this mutant was found by directly sequencing polymerase chain reaction (PCR)-amplified genomic DNA.

(ii) *dpy-10(e128)*. The *dpy-10(e128)* mutation was localized by PCR amplification of the entire *dpy-10* gene from *e128* genomic DNA and subsequent cloning of the products into pBS (Stratagene). Five independent clones were sequenced by using standard dideoxy sequencing and Sequenase version 2.0 (U.S. Biochemical Corp.).

Preparation of RNA. (i) *let-23(n1045)* and *let-23(sy97sy122)*. The strain MT2123/PS48 *let-23(n1045)* was grown at 20°C on special nematode growth factor (NGM) plates up to or just past starvation and was harvested by using standard protocols (53). Nematodes were washed off in EN (0.1 M NaCl, 0.01 M EDTA [pH 7.4]), rinsed twice, cleaned with a sucrose flotation, and frozen in liquid nitrogen. To extract RNA (protocol from Paul Kayne and Russell Hill), the pellet was crushed with a mortar and pestle under liquid nitrogen, resuspended in 3 ml of GTC (4 M guanidinium thiocyanate, 25 mM sodium citrate [pH 7], 0.5% Sarkosyl), and then homogenized with a polytron (generator PT-DA 3007/2). The worm homogenate was layered on top of a CsCl cushion (5.7 M CsCl, 0.1 M EDTA) and centrifuged at 35,000 rpm in an SW55Ti rotor for 15 h. The RNA pellet was resuspended in dimethylpyrocarbonate (DEPC)-treated water, precipitated with LiCl, and resuspended again. RNA from the wild-type strain (courtesy of Jane Mendel) was similarly extracted, except the animals were harvested during exponential growth. RNA from PS463 *let-23(sy97sy122) unc-4(e120); him-5(e1490)* newly starved nematodes was isolated by using the RNazol B method (Tel-Test, Inc.).

(ii) *dpy-10(e128)*. Liquid cultures of either *dpy-10(e128)* or wild-type worms were grown at 20°C in worm growth media and OP50 and harvested before starvation (54). These mixed populations of worms were washed in M9 salts and resuspended in 5 volumes of 5 M guanidine isothiocyanate–10 mM EDTA–50 mM Tris (pH 7.5)–8% β -mercaptoethanol. Worms were homogenized with a polytron for 2 to 3 min and then spun at 10,000 rpm for 30 min at 4°C (Beckman JA-20 rotor). Five volumes of 4 M LiCl were added to the supernatant, which was then allowed to chill at 4°C overnight and then spun as described above. The pellet was resuspended in 5 ml of 3 M LiCl and respun. The pellet was resuspended in 2 ml of 10 mM Tris (pH 7.5)–1 mM EDTA–0.1% sodium dodecyl sulfate and frozen at –20°C. The mixture was allowed to thaw while being vortexed. After a phenol-chloroform extraction, 1/15 volume of 3 M sodium acetate and 2 volumes of ethanol were added to the aqueous phase. RNA was spun down at 10,000 rpm for 20 min at 4°C and resuspended in 0.6 ml of TE (10 mM Tris [pH 8.0], 1 mM EDTA).

Reverse transcription-PCR analysis. (i) *let-23(n1045)* and *let-23(sy97sy122)*. Two micrograms of random hexamers was added to 31 μ g of *n1045* or 13 μ g of wild-type RNA. The tubes were heated to 70°C for 10 min and then placed on ice. Reverse transcription buffer (Bethesda Research Laboratories), deoxynucleoside triphosphates (dNTPs) (final concentration, 0.5 mM each), RNasin (80 units; Promega), and Moloney murine leukemia virus reverse transcriptase (800

units; Bethesda Research Laboratories) were added to a final volume of 80 μ l. The reaction mixtures were incubated at 37°C for 1 h. The above reaction mixture (1.5 μ l) was then PCR amplified (94°C for 3 min; 30 rounds at 94°C for 1 min, at 55°C for 30 s, and at 72°C for 1.5 min; and then a 7-min extension at 72°C after the last round of amplification). The primers CTS5 (exon 16) and CTS3 (exon 18) were used unless otherwise stated (location as shown in Fig. 1A; CTS5 primer sequence: 5'CTCATCTAGGTACAAAACGGAGC C3'; CTS3 primer sequence: 5'CCAAGCTTGTATGAGA TGAATGGCAACGG3'). A negative control was always performed with water as a substitute for the substrate. Reverse transcription of *sy97sy122* RNA and amplification were similarly performed.

To verify that splicing of upstream introns was occurring normally in *n1045*, primers from exons 13 and 15 were used in PCR amplification (data not shown). Only one band with the size expected for a wild-type splice was found. Amplification was also performed with primers from exons 15 and 17 and with primers from exons 15 and 18. The results suggested that intron 15 is also correctly spliced out and that no upstream exon skipping is taking place.

We also performed a control in which reverse transcriptase was left out of the reaction to verify that the *n1045* bands are the result of reverse transcription of RNA (data not shown). Sixty micrograms of DNase-treated RNA was treated as described above for reverse transcription up to the addition of enzyme. The reaction mixture was then split into two tubes. In one tube, 200 units of reverse transcriptase was added; in the other, DEPC-treated water was added. The tubes were incubated at 37°C for 1 h and then PCR amplified. The sample lacking reverse transcriptase produced no bands upon amplification, whereas the sample including reverse transcriptase produced all the expected bands.

(ii) *dpy-10(e128)*. Reverse transcription-PCR analysis was performed as described previously (28). RNA (5 μ g) was reverse transcribed in 1 \times PCR buffer (50 mM KCl, 20 mM Tris [pH 8.4], 2.5 mM MgCl₂, 0.1 mg of bovine serum albumin per ml), 2.5 mM deoxynucleotide triphosphate with 200 U of Moloney murine leukemia virus reverse transcriptase, and 1 pmol of downstream primer D2 at 42°C for 60 min. This reaction mixture (1/20) was used for standard PCR amplification with primers D1 (5'CACATTTTCAAG TGAAGACTTGAG3') and D2 (5'GGTCCTGGAATGCA ACATCCT3'), which produces a fragment of 876 bp on genomic DNA and of 486 bp on normally spliced RNA. The positions of primers used in this study are diagrammed in Fig. 1B. Primer D3 (5'GCGGAATTCATGAGAATAACTA CGAGA3') was used to determine whether exon skipping was occurring in *e128* RNA.

To determine whether the most slowly migrating band in Fig. 4, lane 5, represents a heteroduplex, the two lower bands were cut out of gels, mixed, heated to 95°C for 5 min, and then slowly cooled. Gel electrophoresis of this mixture revealed three bands, the two original bands as well as the third, more slowly migrating species. In addition, the slowest band was excised from a gel, heated to 95°C for 5 min, and then slowly cooled. Gel electrophoresis of this sample resulted in the same three bands.

Sequencing of transcripts. (i) *let-23(n1045)* and *let-23(sy97sy122)*. Reverse-transcribed *n1045* and wild-type RNA were PCR amplified with exon 16 and exon 18 primers. The single band produced by wild-type RNA and the three visible bands produced by *n1045* RNA were subcloned into pBlue-script SK+. Individual colonies were picked, minipreped,

and cut with restriction enzymes to verify insert and size. Most minipreps were sequenced by using a primer from exon 17 (extending toward exon 16) starting at approximately 100 bases from the start of exon 17. Sequencing was performed by using standard double-stranded sequencing reactions and Sequenase version 2.0. Only one strand was sequenced. Point mutations were not detected as a result of PCR amplification. Some subclones, including those lacking exon 17, were sequenced by using vector primers. The wild-type-size band in *sy97sy122* was similarly subcloned. Sequencing was performed by using a primer from exon 18. One of the subclones (described in the text) was over 100 nt shorter than the rest, evidently being a contaminant in the gel-excised wild-type-size band.

(ii) *dpy-10(e128)*. Reverse transcription-PCR products generated by using primers D1 and D2 were cut with *SalI* and *SacI* restriction enzymes, which cut just internal to these primers, yielding a 176-bp fragment (when normal splicing occurs). These fragments were cloned into pBS cut with *SalI* and *SacI*. Clones that contained approximately correctly sized inserts were selected as assayed by restriction analyses and were sequenced by using standard double-stranded dideoxy sequencing with Sequenase version 2.0 and a universal sequencing primer.

Quantitation of RNA. (i) *let-23(n1045)*. Quantitation of the relative abundance of the four *n1045* transcript species was carried out as follows. ^{32}P -end-labeled primers were used to amplify *n1045* cDNA, as described above, for 15, 20, 22, and 25 rounds. The reactions were electrophoresed on a 5% denaturing acrylamide gel. No products were visible at 15 rounds (data not shown). The relative intensities of all four *n1045* bands in each lane were determined by using a PhosphorImager (400S; Molecular Dynamics, Sunnyvale, Calif.). The ratios of the three lower bands to each other in each lane did not vary significantly at 20, 22, and 25 rounds of amplification, suggesting that amplification was not at saturation. Since we are quantitating the ratios of the bands within a given lane, the ratios are internally controlled except for the relative efficiencies of amplification. Since the three lower bands are small and similar in size (264, 372, and 425 bases), we do not expect much error due to inefficiencies in amplification. The relative ratios of the upper band to the others did not vary at 22 or 25 rounds but did vary at 20 rounds, probably because it was the weakest band or because it is much larger (741 bases).

Since one of the bands seen in the autoradiogram was the same size as that expected for genomic DNA, we DNase treated the RNA to determine whether this band was DNA dependent. Sixty micrograms of *n1045* 20°C RNA was treated with 6 U of DNase (Promega) in the presence of 100 U of RNasin (Promega) and buffer (final concentration: 40 mM Tris 7.5, 10 mM NaCl, 6 mM MgCl_2 , 1 mM dithiothreitol) in a total volume of 100 μl . The reaction mixture was incubated at 37°C for 1 h, phenol-chloroform and chloroform extracted, and then precipitated with ammonium acetate. After reverse transcription and PCR amplification with labeled primers, the genomic-size band was still present (data not shown), indicating that this band probably results from lack of splicing at introns 16 and 17 and not from DNA contamination.

RNase protection experiments were performed by using a standard protocol (37). The subclone pIS8A was cut with *Bam*HI and transcribed with T7 RNA polymerase in the presence of $[\alpha\text{-}^{32}\text{P}]\text{UTP}$. The resulting antisense transcript of 385 nt protects 130 nt of *let-23* exon 6. The subclone pIS34 was cut with *AccI* and transcribed with T3 RNA polymerase.

This 140-nt antisense transcript protects 100 nt of *let-23* exon 15, just upstream of the tyrosine kinase domain. Both these probes should be common to all *let-23(n1045)* transcripts. The subclone pT7/T3-18-103 (courtesy of M. Krause) was cut with *Eco*RI and transcribed with T3 polymerase. This 250-nt antisense transcript protects 214 nt of the 3' untranslated region of the actin gene *act-1*. The actin probe was transcribed at one-third the specific activity of the *let-23* probes. All probes were gel purified prior to hybridization. For the two *let-23* probes, 5×10^9 radiolabeled probes were hybridized to approximately 20 μg of *n1045* or wild-type total RNA. For the actin probe, 6×10^9 radiolabeled probes were hybridized to 2 μg of *n1045* or wild-type total RNA. Controls were carried out for each probe with yeast tRNA (30 μg) in place of *C. elegans* RNA. The remainder of the protocol was carried out as described by Miner and Wold (37) with RNases T1 and A. The PhosphorImager counts detected in each protected band for *let-23* RNA, wild-type RNA, and tRNA, respectively, were 3.4×10^7 , 7.2×10^7 , and 5.3×10^5 with the actin probe; 3.6×10^6 , 3.2×10^6 , and 3.4×10^5 with the pIS8A probe; and 3.9×10^6 , 2.8×10^6 , and 3.4×10^5 with the pIS34 probe. After subtracting out noise (tRNA lanes) and normalizing for the actin probe, the pIS8A and pIS34 protections, respectively, suggest that *let-23* transcript levels in *let-23(n1045)* are elevated to 2.4 and 3.2 times the levels in the wild type.

(ii) *dpy-10(e128)*. To determine levels of *dpy-10* message in *e128* RNA, competitive quantitative PCR was used (18, 19). This method used a known amount of competitive template which is coamplified in PCR along with the unknown quantity of cDNA from reverse-transcribed *e128* RNA. The competitive template used was a genomic DNA clone of the *dpy-10* gene, pAL5. Parallel samples, each containing 30 ng of reverse-transcribed *e128* RNA and serially diluted amounts of pAL5 (ranging from 0.01 to 1.00 amol), were amplified by using D4 (5'-GATCTACCGGTGTGTCAC-3') and D2. Amplification with D4 and D2 of unspliced mRNA or genomic DNA yields a 268-bp fragment; amplification of spliced mRNA yields a 220-bp fragment. At the point where the cDNA and genomic DNA templates are amplified equally, the starting concentration of cDNA equals the concentration of the pAL5 template. However, since some of the *dpy-10* transcripts are unspliced for intron 2, these transcripts yield a product that is the same size as the competitor, making it difficult to measure this point of equal intensity; also, for this reason, the amplified signal of the *e128* RNA-cDNA is always lower than that of the wild type. We therefore took as the measure of equal concentrations the point at which the concentration of competitor DNA decreased the original RNA (cDNA) signal to one-half of its value with no competitor added. PCRs were carried out with 1.0 μl of $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ (>400 Ci/mmol), and the radioactivity in each band was then quantified with a Fujix BAS2000 Bio-Image Analyzer which determines radioactivity in units of phospho-stimulable luminescence. The same procedure was done with wild-type RNA. To control for differences in RNA preparations, the same competition was performed with the two RNAs by using *sqt-1*-specific primers. The *sqt-1* gene is a cuticle collagen, which is similar in structure to *dpy-10* (33). Its expression is presumably not affected by mutations in *dpy-10*.

RESULTS

***let-23(n1045)* and *dpy-10(e128)* 3' splice site mutations.** The product of the *C. elegans* gene *let-23* is similar in primary

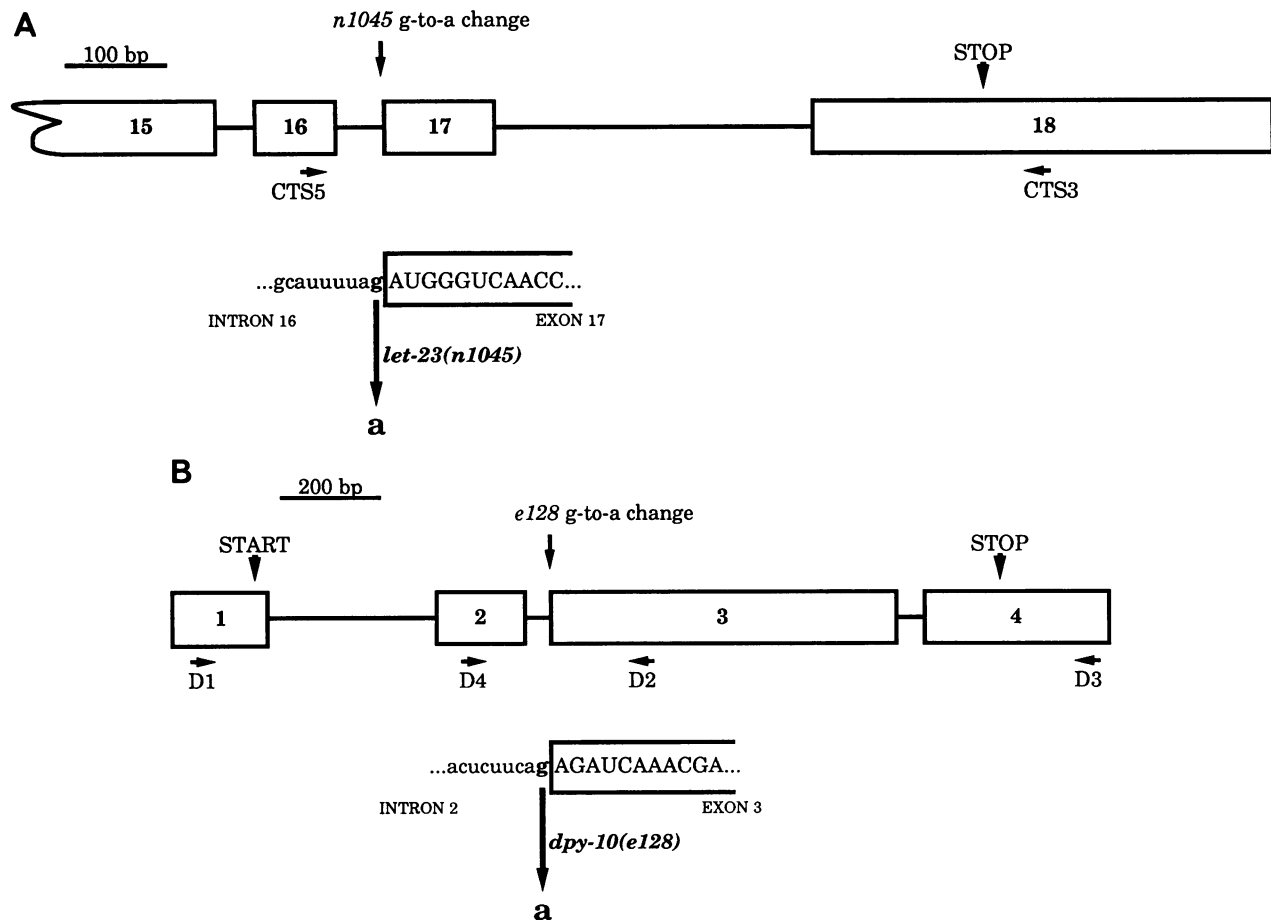


FIG. 1. (A) Location of *let-23(n1045)* point mutation. The exon-intron structure of the C terminus of the *let-23* gene is shown. There are 18 exons in *let-23*, but only the portion from part of the way through exon 15 to the 3' untranslated region is shown. The gene extends for over 7 kb upstream of the region shown. The kinase domain is just 5' to the region shown. Exons are shown as boxes; introns are shown as solid lines between the boxes. The number in each box represents the exon number. The location of the *n1045* mutation (G to A) in intron 16 is indicated above the schematic. Below the schematic, the sequence at and around the mutation is shown. Uppercase letters designate the exon sequence, and lowercase letters designate the intron sequence. The locations of primers CTS5 and CTS3 used to analyze *n1045* transcripts are indicated. (B) Location of *dpy-10(e128)* mutation. The entire *dpy-10* gene is diagrammed similarly to that for panel A. The location of the *e128* intronic mutation (AG to AA) is indicated above the schematic, as are the locations of the primers used to analyze the *e128* transcripts. Below the schematic, the sequence at and around the *e128* mutation is shown. START and STOP refer, respectively, to the locations of the start and stop of translation.

sequence to the mammalian epidermal growth factor receptor (2). Genetic studies have shown that the *let-23* receptor tyrosine kinase is required for viability past the first larval stage, for hermaphrodite fertility, and for the development of the posterior ectoderm, the hermaphrodite vulva, and the male spicules (3, 13). Elimination of *let-23* activity results in completely penetrant larval lethality (3).

The *dpy-10* gene is structurally and functionally quite different: *dpy-10* encodes a cuticular collagen (35a) that is similar in structure to the other sequenced cuticle collagens (31–33, 60). Mutations in the *dpy-10* collagen gene often result in a dumpy phenotype such that the mutant nematodes are shorter and fatter than the wild type (7). In addition, mutations in the *dpy-10* gene can sometimes result in a roller phenotype—the nematodes roll while they move because of a helical twist of the cuticle and underlying structures (11). Elimination of the *dpy-10* collagen gene results in nematodes which are both dumpy and left-handed rollers (35a).

The *let-23(n1045)* mutation was localized and sequenced

as part of a structure-function study of *let-23* (2a). Many alleles of *let-23* have been isolated and genetically characterized (3, 13, 14). The non-null, reduction-of-function allele *n1045* is particularly interesting because it appears to be defective in negative regulation of vulval development (3). The *n1045* allele was sequenced (see Materials and Methods), and it was found that the last base of intron 16 is altered from G to A (Fig. 1A). This splicing mutation eliminates the invariant AG at the 3' splice site.

A number of *dpy-10* alleles, including *dpy-10(e128)*, were also sequenced as part of a structure-function study. The allele *dpy-10(e128)* reduces but does not eliminate *dpy-10* function. Sequencing of the *e128* allele revealed that the last nucleotide of intron 2 is altered from G to A (Fig. 1B), thus changing the AG dinucleotide to AA. No other nucleotide change is present within the *dpy-10* gene.

Although the studies were carried out independently, we are reporting the studies of *let-23(n1045)* and *dpy-10(e128)* together to reinforce the results and conclusions with respect

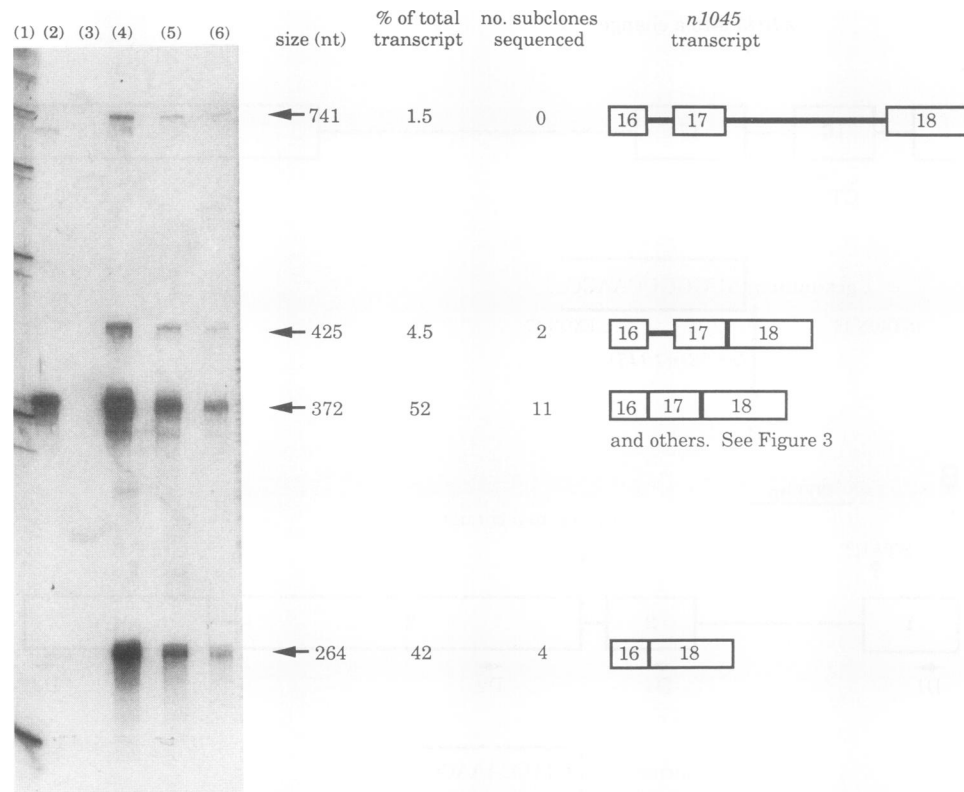


FIG. 2. Reverse transcription-PCR analysis of wild-type and *n1045* RNA. PCR amplification of wild-type and *n1045* reverse-transcribed RNA was done with radiolabeled CTS5 and CTS3 primers. The products were electrophoresed on a 5% denaturing acrylamide gel, and the gel was autoradiogrammed. Lanes: 1, 123-base ladder; 2, amplification of wild-type cDNA for 25 rounds; 3, negative PCR control (no substrate). Amplification of *n1045* cDNA was done for 25 (lane 4), 22 (lane 5), and 20 (lane 6) rounds. To the right, the sizes of the bands in the *n1045* lane, the percentages that each *n1045* band represents relative to all the bands in the *n1045* lane, the numbers of subclones from each band sequenced, and the exon-intron structure of *n1045* transcripts found in each band are indicated (exon-intron structure is given as shown in Fig. 1A). The exon-intron structure of the upper band is based on size only and not on sequence data. Seven subclones from the wild-type band were also sequenced as a control. There is a faint extra band in the wild-type lane. Since it is not at the size predicted for any splicing product, it could be caused by non-*let-23*-specific amplification. The weaker band immediately below the *n1045* band at the size expected for a wild-type splice is probably the result of cryptic splicing at the AG 17 bases downstream of the normal 3' splice site.

to splicing. The significance of these results with respect to the function of the *let-23* and *dpy-10* genes will be discussed separately (2a, 35a).

Splicing in *let-23(n1045)*. To ascertain the effect of the *n1045* mutation on *let-23* mRNA, we isolated total RNA from homozygous *let-23(n1045)* hermaphrodites, reverse transcribed the RNA by using random primers, and then performed PCR amplification with primers from exon 16 and the 3' untranslated region in exon 18. The results are shown in Fig. 2. Amplification of reverse-transcribed RNA from wild-type hermaphrodites yields only one band at the expected size of 372 nt (exon 16-exon 17-exon 18). However, amplification of reverse-transcribed RNA from *n1045* yields four bands. Two of these bands are consistent in size with the effects of 3' splice site mutations in other systems: exon skipping (exon 16-exon 18) and a failure to splice (exon 16-intron 16-exon 17-exon 18). The other two bands are unexpected: a prominent band at the size expected for a wild-type splice (exon 16-exon 17-exon 18) and a fainter band at the size expected for two unspliced introns (exon 16-intron 16-exon 17-intron 17-exon 18). The relative intensities of these four bands were determined and represent, respectively, 1.5% (the exon 16-intron 16-exon 17-intron 17-exon 18 band), 4.5% (the exon 16-intron 16-exon 17-exon

18 band), 52% (the exon 16-exon 17-exon 18-size band), and 42% (the exon 16-exon 18 band) of the total transcript species in *n1045*.

Sequencing verified that the lower band at 264 nt contained transcripts generated by exon skipping (exon 16-exon 18) and that the band at 425 nt, slightly above that of the wild type, contained transcripts with unspliced intron 16 (exon 16-intron 16-exon 17-exon 18). The upper band at 741 nt was not sequenced since it was too difficult to cut out of an ethidium-stained gel. Nonetheless, the size is as predicted for unspliced introns 16 and 17 (exon 16-intron 16-exon 17-intron 17-exon 18).

We were surprised at the results of sequencing 11 subclones from the *n1045* band at the size (372 nt) expected for a wild-type splice (Fig. 3). In two subclones, a cryptic AG, 17 bases downstream of the mutated intron 16 3' splice site, was used as a 3' splice site. However, in five subclones, a splice had occurred where the splice occurs in the wild-type despite the fact that there was an AA and not an AG 3' acceptor. In the remaining four subclones, a cryptic AU (two different ones), GG, and UG were used as the 3' splice site. These four non-AG cryptic sites are all within 8 bases of the normal 3' splice site. Despite the great variation of 3' splice sites used, the normal 5' splice donor site of intron 16 was

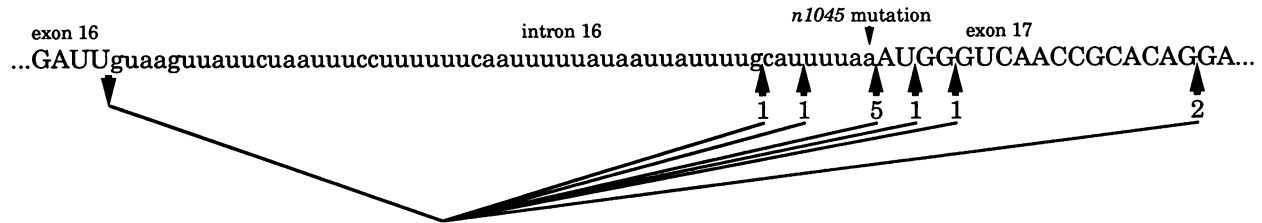


FIG. 3. Splicing near the intron 16 3' splice site in *let-23(n1045)*. Eleven subclones from the *n1045* band at the size expected for a wild-type splice product were sequenced. The splices inferred from that sequencing are shown. Lowercase letters denote the intron sequence; uppercase letters denote the exon sequence. The larger downward-pointing arrowhead indicates the 5' junction used for all splices and is the correct 5' splice site. The upward-pointing arrowheads indicate the 3' splice junctions used. The numbers beneath each arrowhead indicate the number of subclasses in which a given 3' splice site was found.

used for all spliced transcripts. It is possible that we have not identified all the transcripts found in *let-23(n1045)* since some of the splice products are represented by single subclones.

As shown in Fig. 2, reverse transcription-PCR analysis of wild-type RNA results in only one band at the expected size. As a control, we sequenced seven subclones from this band. All indicate a perfect splice from exon 16 to exon 17. Thus, the unusual *n1045* splicing products do not occur in the wild type and are unlikely to be an artifact of PCR.

We performed several other controls (details in Materials and Methods). First, the results are reverse transcriptase dependent. Second, splicing of upstream introns occurs normally. Third, none of the *n1045* bands disappear upon treatment of the RNA with DNase prior to reverse transcription, indicating that the bands seen, in particular the unspliced exon 16-intron 16-exon 17-intron 17-exon 18 transcript, are the result of RNA and not DNA. Lastly, two other independent *n1045* RNA preparations result in the same bands after reverse transcription-PCR amplification.

We quantitated the abundance of *let-23* message in the *n1045* mutant by RNase protection. We used two different *let-23* antisense probes that are common to all *let-23* transcripts and normalized to an actin antisense probe. The data indicate that the combined level of all *let-23* transcripts in the *n1045* mutant is elevated to approximately threefold that of the single *let-23* message in the wild type. This result was verified with quantitative and competitive PCR experiments, which indicate that relative to the *rol-6* collagen gene, the combined level of all *let-23* messages in *n1045* is elevated to roughly four- to fivefold that of the single *let-23* message in the wild type (data not shown). A caveat to interpreting this result is that the physiologies of the *n1045* and wild-type strains might differ (e.g., there are many dead larvae associated with *n1045*). Nonetheless, we conclude that the amount of *let-23* transcript in *n1045* is significant relative to the amount of *let-23* transcript in the wild type.

Splicing in *dpy-10(e128)*. To determine the splicing behavior of the *dpy-10(e128)* mutant, reverse transcription and PCR amplification were performed with RNA isolated from *e128* animals. The results are shown in Fig. 4. Amplification of reverse-transcribed wild-type RNA results in a single 486-bp band, as expected for normal splicing of introns 1 and 2 (lane 4). Amplification of reverse-transcribed *e128* RNA results in three bands (lane 5). The predominant band, representing about two-thirds of the total signal, appears at the size of the normally spliced product. A larger, 534-bp band appears at the size expected when intron 1 is normally spliced but intron 2 is unspliced. We have determined that a third band, which migrates more slowly than the other two

bands, is a heteroduplex consisting of one strand of unspliced product and one strand of spliced product. Examples of such heteroduplexes have been previously reported (61, 64). To determine whether exon skipping was occurring, primer D3 (Fig. 1B) was used for reverse transcription-PCR analysis. No product that would suggest that exon 3 is skipped in any *e128* transcripts was detected (data not shown).

The reverse transcription-PCR products in the apparently normally spliced band at 486 bp were cloned, and 13 subclones were sequenced. Three distinct 3' splice sites were found to be used (Fig. 5). In two clones, splicing occurred at the same location as in the wild type, despite the presence of the AA dinucleotide; in four clones, splicing occurred at an upstream AA acceptor at position -8 relative to the location

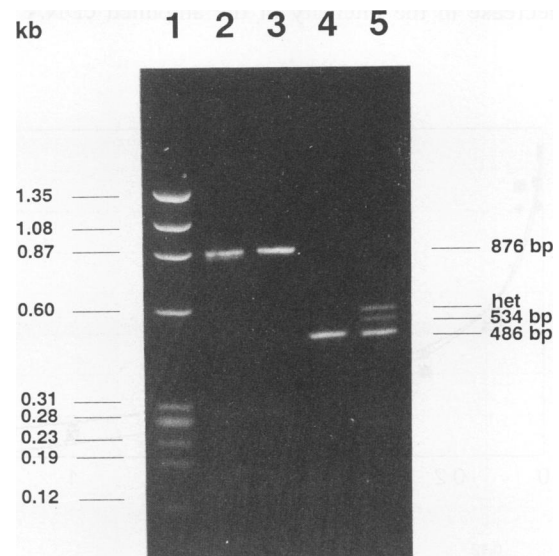


FIG. 4. Reverse transcription-PCR analysis of wild-type and *e128* RNA. PCR amplification for 30 rounds of wild-type and *e128* reverse-transcribed RNA was done with primers D1 and D2 (see Fig. 1B). Lanes: 1, ϕ X174 marker cut with *Hae*III; 2, amplification of wild-type DNA; 3, amplification of *dpy-10(e128)* DNA; 4, amplification of wild-type cDNA; 5, amplification of *e128* cDNA. The sizes of genomic DNA (876 bp of 5' untranslated exon 1-intron 1-exon 2-intron 2-exon 3), unexcised intron 2 (534 bp of 5' untranslated exon 1-exon 2-intron 2-exon 3), and spliced RNA (486 bp of 5' untranslated exon 1-exon 2-exon 3) are indicated on the right, as is the spliced-unspliced heteroduplex (het) (see Materials and Methods).

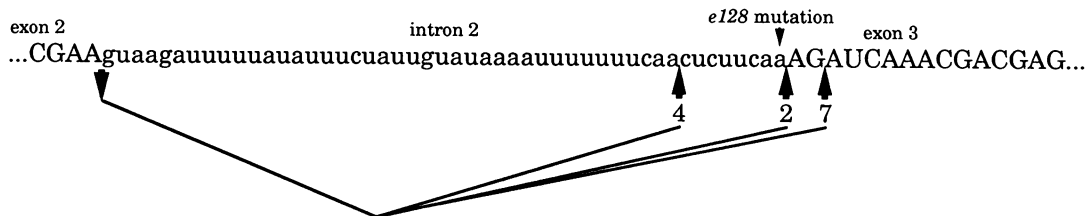


FIG. 5. Splicing near the intron 2 3' splice site in *dpy-10(e128)*. Thirteen subclones from the 486-nt *e128* band (Fig. 4, lane 5) were sequenced. The splices inferred from the sequencing are shown. The results are represented as described in the legend to Fig. 3.

of the wild-type 3' splice site; in the remaining seven, splicing occurred at the AG that immediately follows the mutated site, corresponding to position +2. As a control, two wild-type *dpy-10* cDNAs were sequenced and found to splice at the expected 3' acceptor. Thus, consistent with the results for *let-23(n1045)*, the splicing machinery can still recognize and use the mutant splice site at least some of the time despite the absence of the universally conserved AG dinucleotide. Also surprising is the utilization of the acceptor at -8 that also consists of an AA rather than an AG dinucleotide.

To determine the level of *dpy-10* transcripts in *e128* RNA, we performed quantitative, competitive PCR (see Materials and Methods) (18, 19). We performed these experiments with *e128* and wild-type RNAs to determine whether a difference in *dpy-10* transcript levels could be detected. To control for differences in the RNA preparations, competitive PCR was also carried out with *e128* and wild-type RNAs by using primers from the *sqt-1* collagen gene (33).

The results are shown in Fig. 6, in which we have charted the decrease in the intensity of the amplified cDNA band

(ordinate) as increasing amounts of competitor DNA are added (abscissa). The point at which the cDNA product decreases to one-half of its maximum signal (with no competitor added) represents the point at which the initial concentrations of DNA competitor and cDNA are equal. Figure 6A shows that the wild-type and *e128* RNA preparations give nearly identical results when the *sqt-1* control cDNA is amplified, indicating that the levels of *sqt-1* in the two RNAs are very similar. Thus, differences seen when the *dpy-10* cDNAs are amplified are probably not due to differences between the two RNA preparations. Figure 6B shows the results of the *dpy-10* competition. The point at which the cDNA band decreases to one-half of its original signal is at 0.07 and 0.11 amol of competitor DNA for *e128* and wild-type RNA, respectively. We therefore estimate that the total level of *dpy-10* message in *e128* RNA is 65% of that in the wild type. Thus, as with *let-23(n1045)*, the unusual splicing products seen are occurring at significant levels relative to wild-type *dpy-10* splicing products.

Splicing in *let-23(sy97sy122)*. To determine whether our results applied to longer *C. elegans* introns, we analyzed

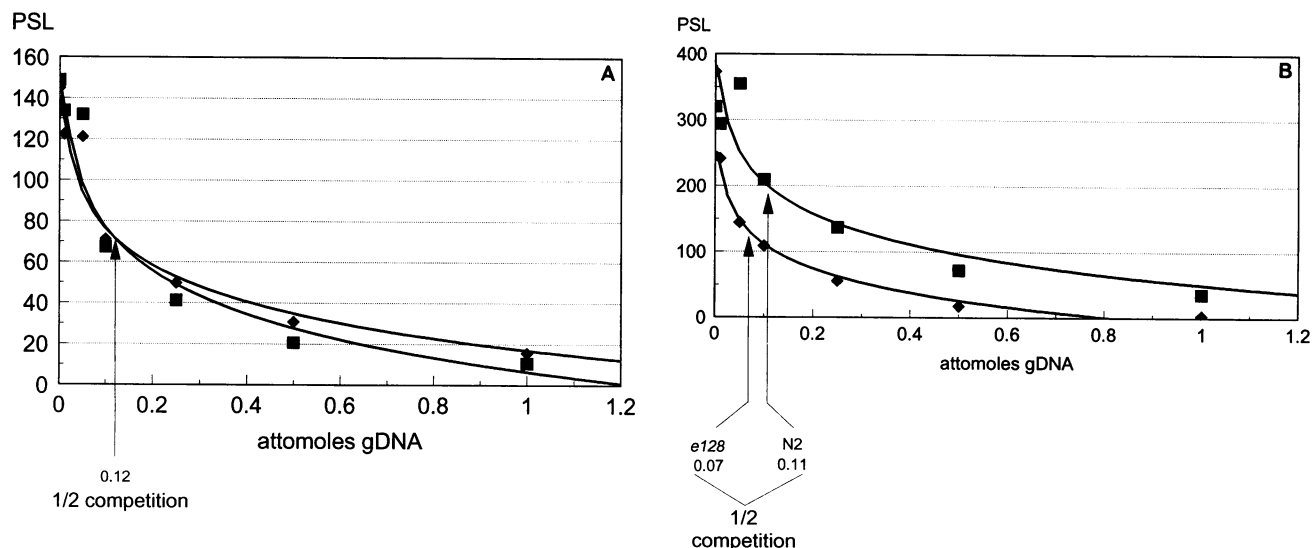


FIG. 6. Quantitation of *dpy-10* transcript levels in wild type and *dpy-10(e128)*. Competitive quantitative PCR was performed as described in the text. (A) Decrease in signal of *sqt-1* control cDNA when increasing amounts of competitor DNA (a plasmid clone) are added to the PCR mixture. The point at which the signal decreases to one-half of its maximum represents the point at which the initial concentrations of cDNA and competitor are equivalent. (B) Results of *dpy-10* quantitation with primers D4 and D2. The level of cDNA in *e128* is lower at all points than that of the wild type because a fraction of the *dpy-10* cDNAs in *e128* is unspliced and therefore runs at the same position as the competitor. Nonetheless, the point at which one-half competition is reached is unaffected. The estimate of the amount of *dpy-10* transcript in the wild-type RNA is 0.11 amol, while in *e128* RNA it is 0.07 amol. The competitor used is a plasmid clone containing the wild-type *dpy-10* gene. Quantitation was performed by including a radiolabeled nucleotide in the PCR mix and the use of a phospho-imager. ■, N2 RNA; ♦, *e128* RNA; PSL, phospho-stimulable luminescence.

distribution at 45 to 55 nt (6, 15). In contrast, most introns of other higher eukaryotes are 75 to 2,000 nt long, and efficient splicing of these introns may require a minimal length of approximately 80 nt (63). Hence, there may be elements of the *C. elegans* splicing apparatus which differ from those of higher eukaryotes to efficiently splice such a large percentage of small introns. Another invertebrate, *Drosophila melanogaster*, also has short introns (24).

There are other unusual features of splicing in *C. elegans*. Many *C. elegans* introns lack a potential branch point sequence (pyrimidine-purine-A-pyrimidine or YRAY) such as is found in higher organisms (6, 15). In addition, *C. elegans* introns have a highly conserved 3' splice site: UUUCAG, with a U present at -5 relative to the cleavage site in over 98% of the introns (6, 15). Lastly, *C. elegans* introns, like plant introns, have an elevated A-U content just upstream of the 3' splice site that may be important for their processing (5, 62). Conversely, just upstream of the 3' splice site of higher eukaryotic introns is the polypyrimidine stretch that is important for their splicing (44, 45, 48, 51).

To determine whether *C. elegans* introns may be spliced differently from mammalian introns, the splicing of *C. elegans* short introns (about 50 nt) was assayed in mouse cells and HeLa cell extracts (29, 42). In neither case did correct splicing occur. In only one of these two cases (42), correct splicing was restored by lengthening the intron.

Intron 16 of *let-23* and intron 2 of *dpy-10* fit the above profile of *C. elegans* introns. They are both short (53 and 48 nt, respectively), relatively A-U rich, and fit the 3' splice site consensus sequence in five of six places. There are three potential branch point sequences (YRAY) in intron 16 of *let-23* (located at -19, -28, and -43) and one in intron 2 of *dpy-10* (located at -11). The YRAY sequence in *dpy-10* is closer to the 3' splice site than is usual for a branch point sequence.

Our *in vivo* results indicate that for these short *C. elegans* introns, an AG is important, but not essential, for the first step of splicing, i.e., cleavage at an upstream 5' splice site. In both *let-23(n1045)* and *dpy-10(e128)* [the mutation *let-23(sy97sy122)* is discussed below], mutation of the AG results in some stable transcripts with an unspliced intron, indicating that the AG does influence the efficiency of the first step of splicing. However, we cannot exclude that this effect is indirect, e.g., the mutation at the 3' splice site could speed up the transport pathway. Nonetheless, splicing can occur without the AG, and, when it does, the correct 5' splice site is chosen. Thus, to the extent analyzed, the downstream AG does not influence the accuracy of cleavage at the 5' splice site.

Our results also indicate that for these introns AG is important, but not essential, for the second step of splicing, i.e., cleavage at the 3' splice site and exon ligation. In both *let-23(n1045)* and *dpy-10(e128)*, splicing to the correct site occurs even though the splice acceptor was mutated from AG to AA. From this, and other non-AG splices seen, we can also conclude that an AG per se is not essential for cleavage and ligation at the 3' splice junction, although our data suggest that there must be a match to either the A or the G. This preference for an A in the penultimate position or a G in the ultimate position may reflect some base pairing interaction with a small nuclear RNA (e.g., U1) (20 [and references therein]). Despite the unusual splices that still occur when the AG is mutated, the data do indicate that an appropriately placed AG is important for proper splicing since in wild-type animals only the proper splice occurs.

A model for 3' splice site selection in short *C. elegans*

introns. What then determines the pattern of promiscuous splicing seen close to the wild-type 3' splice sites of introns 16 and 2, respectively, of *let-23(n1045)* and *dpy-10(e128)*? The *C. elegans* splicing machinery may derive intron-exon border information based on the AU versus the GC content, as has been proposed for splicing in plants (5, 62). In both the wild type and the mutants, this information could position the machinery in the proper vicinity of the 3' splice site. Unlike the wild type, however, the site of subsequent cleavage and ligation is not perfectly determined.

Analysis of the actual splices made (Table 1) suggests that the exact location of the splices in the mutants may depend on competition among several of the sequence elements discussed above: (i) an AG (i.e., any AG), (ii) the remainder of the *C. elegans* 3' splice site consensus (i.e., UUUC, with the highly conserved U underlined), (iii) an A-U-rich region, and (iv) a putative branch site sequence. For example, of the splices made in *dpy-10(e128)*, all are downstream of an A-U-rich region. However, the non-AG splices made do have a U at position -5, whereas the AG(+2) splice does not. This contrasts with the wild-type case in which the splice occurs at an AG with a U at -5. On the other hand, the AG(+15) is not used, perhaps because it cannot compete with the other AG(+2) and because there is neither a U at -5 relative to this AG nor an A-U-rich region immediately upstream. For *e128*, there is no obvious role of potential branch point sequences in 3' splice selection. For *let-23(n1045)*, the analysis is similar. In all cases but the GG(+4) splice, a U exists at -5 relative to a non-AG splice. Furthermore, there is no U at -5 relative to the splice at the cryptic AG, and this splice is the only one not preceded by an A-U-rich region. As with *e128*, it may be that the cryptic AG can sometimes be ignored if other nearby dinucleotides contain other sequence elements important for splicing. Unlike all the splices made in the mutant, all four sequence elements are present in the wild-type gene. The GG(+4) splice is perhaps made without a U at -5 because, after AG, GG might be the best dinucleotide at the 3' splice acceptor site (12).

We hypothesize that once the machinery is positioned at the 3' splice site, the location of cleavage and ligation is determined by an AG, an upstream motif related to UUUC, the A-U-rich region itself, and, perhaps to a lesser extent, the presence of a suitable branch point sequence. In the wild-type *let-23* and *dpy-10* introns studied here, these preferences all cooperate so that splicing occurs at the proper position. When the normal AG is mutated, these preferences allow other potential sites to compete with each other for selection as the 3' splice site. Consistent with the notion that the four elements noted together contribute to delineate a 3' splice site, our data also indicate that no one element alone is absolutely required for splicing. We cannot discern any influence in the selection of the 3' splice sites by adjacent exonic nucleotides, as has been suggested for yeast cells (41).

The presence of unspliced intron in the mutants but not in the wild type may indicate that in the mutants either a critical, timely recognition of the 3' splice site did not occur and these unspliced transcripts have escaped splicing altogether or the timing of splicing machinery is significantly slowed down and these unspliced transcripts have yet to be spliced. The presence of significant levels of exon skipping in *n1045* suggests that the intron 17 3' splice site can efficiently compete with the mutated intron 16 3' splice region for splicing to the intron 16 5' splice donor site. Exon skipping may not be occurring with *e128* because the size of the intron

TABLE 1. Characteristics of different 3' splice sites used in *dpy-10(e128)*, *let-23(n1045)*, and *let-23(sy97sy122)* transcripts

Mutation	Dinucleotide at splice (location) ^a	No. of times sited used/ Total no. analyzed	Occurrence of AG	Adjacent upstream sequence ^b	A-U rich ^c	BPS ^d
<i>dpy-10(e128)</i>	AG (+2)	7/13	Yes	UCAA	Yes	+
	AA (-8)	4/13	No	U <u>U</u> UC	Yes	-
	AA (0)	2/13	No	C <u>U</u> UC	Yes	+
	AG (+15)	0/13 ^e	Yes	GACG	No	++
<i>let-23(n1045)</i>	AA (0)	5/11	No	U <u>U</u> UU	Yes	++
	AG (+17)	2/11	Yes	GCAC	No	++
	UG (-8)	1/11	No	A <u>U</u> UU	Yes	++
	AU (-5)	1/11	No	U <u>U</u> GC	Yes	++
	AU (+2)	1/11	No	U <u>U</u> AA	Yes	++
	GG (+4)	1/11	No	AAA <u>U</u>	Yes	++
	AG ^f	4/4	Yes	U <u>U</u> UC	Yes	++
<i>let-23(sy97sy122)</i>	AA (0)	10/10	No	U <u>U</u> UC	Yes	++
	AG (+190)	1/1 ^g	Yes	U <u>U</u> UC	No	++

^a Dinucleotide at the 3' splice site used for various splices; in parentheses, the locations of the splice made relative to the location of the wild-type splice site (-, upstream; +, downstream) are indicated.

^b The sequence of the four bases upstream of the dinucleotides adjacent to the splice site. The consensus sequence in *C. elegans* is UUUC. The highly conserved U is underlined.

^c Yes indicates that the region upstream of the splice is A-U rich; no indicates that the region is not A-U rich.

^d ++ indicates a well positioned branch point sequence (BPS) 18 to 40 nt upstream of the 3' cleavage site; + indicates a poorly positioned BPS; - indicates no nearby BPS.

^e This site has not been shown to be used. It is included here to emphasize that, although it is a physically well positioned AG, it competes poorly with the two non-AG 3' splice sites.

^f This splice is made when the 5' splice donor of intron 16 is paired with the 3' splice acceptor of intron 17.

^g See Materials and Methods for isolation of this subclone.

that needs to be spliced out for exon skipping to occur is longer than in *n1045* (799 versus 470 nt; the 3' acceptor at intron 3 of *e128* is UUCAG).

3' Splice site selection in longer *C. elegans* introns. The introns mutated in *let-23(n1045)* and *dpy-10(e128)* are atypically short for most eukaryotes. Our analysis of mutations in the longer intron 17 (316 nt) of *let-23* suggests that some, but not all, of our results with the short introns are applicable to long introns. The 3' splice site of intron 17 matches the consensus in six of six places and YRAY sequences exist at -14 and -40 (Table 1).

Mutation of the intron 17 3' splice acceptor from AG to AA results in splicing exclusively to the adjacent exonic AG (2a). Thus, in contrast to the results with *dpy-10(e128)*, the mutated AA cannot compete with an adjacent AG despite the presence of an otherwise perfect consensus. Furthermore, in both this single mutant and the double mutant *sy97sy122*, only one rather than several 3' splice sites is utilized. Together these results suggest that the selection of 3' splice sites for longer *C. elegans* introns may be less flexible and somewhat different from that for short introns. Along these lines, it has been suggested that the informational contents of short and long *C. elegans* introns are different (15).

Nonetheless, splicing in the double mutant *sy97sy122* indicates that an AG is not required for splicing, even of a long *C. elegans* intron. Splicing occurs primarily at the normal (now AA) 3' splice site despite the presence of an AG 25 nt upstream. Thus, as with short introns, elements other than an AG can direct the splicing of long introns. Although we have not quantitated RNA levels in *sy97sy122*, the fact that most of the *let-23* phenotypes are close to those of the wild type in this mutant (27a) suggests that significant levels of splicing to the normal site are occurring. The presence of the splice at the UUUCAG 190 nt downstream of the normal site (Table 1) supports our model for 3' splice selection: to

splice at this site, seven other AGs, none of which have a perfect consensus, must be passed over.

Implications for splicing in other organisms. Although our results may be a consequence of some unique aspect of splicing in *C. elegans* and therefore limited to *C. elegans*, there is some precedent for our results in yeasts and possibly mammals. Mutation of the AG at the 3' end of the yeast *Saccharomyces cerevisiae* actin intron (AG to GG or AC) results in low levels of in vitro and in vivo splicing at the correct (but mutated) 3' splice site (16, 59). In the actin intron of the yeast *Kluyveromyces lactis*, a GG can compete with the normal AG as the 3' splice site (12). Furthermore, in one study of human analbuminemia, although the AG at a 3' splice site was mutated to a GG, very low levels of albumin were still detected in the bloodstream, suggesting that trace amounts of correctly spliced transcripts were made (47). In a study of steroid 21-hydroxylase deficiency in humans, S1 nuclease analysis of RNA suggested that a mutation which created a 3' splice site (AC to AG) resulted in three abnormal, closely spaced splicing products that use AG and AC as 3' splice sites (26). It is also possible that results similar to ours have been missed in the past. Some previous studies have relied on S1 nuclease and/or gel analysis assays, which may be less sensitive than sequencing of PCR-amplified cDNAs. Thus, our results may not be unique to *C. elegans*.

As discussed above, our data also suggest that primary sequences apart from the AG and branch point sequence influence 3' splice site selection and that these primary sequences may have an incremental effect on defining the competitiveness of 3' splice sites. Consistent with our findings, there is growing evidence that elements besides the AG and branch point sequence (e.g., other primary sequences, secondary structures, and 5' splice sites) may influence the strength and selection of 3' splice sites (12, 40, 43, 44, 46, 63).

The presence of exon skipping (exon 16-exon 18) and the

occasional inclusion of intron 17 (exon 16-intron 16-exon 17-intron 17-exon 18) in *n1045* RNAs suggest that the intron 16 mutation is affecting splicing of the downstream intron. These results, especially the latter, are consistent with the idea that splicing occurs in a unitary complex of multiple introns (38) or that an early step of splicing is exon definition (46). According to this latter model, the inefficient association of the splicing machinery at the 3' splice site of intron 16 may affect the ability of the splicing apparatus to associate with the 5' splice site of intron 17.

Finally, the findings of this study suggest a mechanism for generating diversity in mRNA transcripts—competition among many closely positioned 3' splice sites, some of which may not be AG. There are two relevant studies along these lines. One is mentioned above (26). In the other, analyses of rat $\alpha 2\mu$ globulin mRNAs suggest that transcript diversity for this family of genes may be generated in part by using closely positioned AGs and UGs as 3' splice sites (17).

ACKNOWLEDGMENTS

We thank Gregg Jongeward for the mutant *let-23(sy97sy122)*, Tom Blumenthal for discussions on splicing in *C. elegans*, and Bob Horvitz for suggesting promiscuous splicing as a mechanism for generating diversity. We thank James Deshler, Andy Golden, Paul Kayne, and John Rossi for discussions and comments. We thank Jeff Miner for help with RNase protection and Sean Tavtigian and the Dervan Laboratory for help with PhosphorImager quantitation. We thank Jane Mendel for wild-type RNA, Mike Krause for plasmid pT7/T3-18-103, Joseph LaRusso for help with sequencing, and Paul Kayne, Russell Hill, and Giovanni Lesa for technical advice.

P.W.S. is an investigator of the Howard Hughes Medical Institute. This research was supported by grants HD23690 to P.W.S. and HD22028 to J.M.K.

REFERENCES

- Aebi, M., H. Hornig, R. A. Padgett, J. Reiser, and C. Weisman. 1986. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 47:555–565.
- Aroian, R. V., M. Koga, J. E. Mendel, Y. Ohshima, and P. W. Sternberg. 1990. The *let-23* gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. *Nature (London)* 348:693–699.
- Aroian, R. V., G. Lesa, and P. W. Sternberg. Unpublished data.
- Aroian, R. V., and P. W. Sternberg. 1991. Multiple functions of *let-23*, a *Caenorhabditis elegans* receptor tyrosine kinase gene required for vulval induction. *Genetics* 128:251–267.
- Atweh, G. F., N. P. Anagnou, J. Shearin, B. G. Forget, and R. E. Kaufman. 1985. β -Thalassemia resulting from a single nucleotide substitution in an acceptor site. *Nucleic Acids Res.* 13:777–790.
- Blumenthal, T. 1992. Personal communication.
- Blumenthal, T., and J. Thomas. 1988. *cis* and *trans* mRNA splicing in *C. elegans*. *Trends Genet.* 4:305–308.
- Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* 77:71–94.
- Carstens, R. P., W. A. Fenton, and L. R. Rosenberg. 1991. Identification of RNA splicing errors resulting in human ornithine transcarbamylase deficiency. *Am. J. Hum. Genet.* 48:1105–1114.
- Chen, S.-H., M. Zhang, A. R. Thompson, G. L. Bray, and C. R. Scott. 1991. Splice junction mutations in factor IX gene resulting in severe hemophilia B. *Nucleic Acids Res.* 19:1172.
- Cladaras, C., M. Hadzopoulou-Cladaras, B. K. Felber, G. Pavlakakis, and V. I. Zannis. 1987. The molecular basis of a familial apoE deficiency. *J. Biol. Chem.* 262:2310–2315.
- Cox, G. N., J. S. Laufer, M. Kusch, and R. S. Edgar. 1980. Genetic and phenotypic characterization of roller mutants of *Caenorhabditis elegans*. *Genetics* 95:317–339.
- Deshler, J. O., and J. J. Rossi. 1991. Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev.* 5:1252–1263.
- Ferguson, E. L., and H. R. Horvitz. 1985. Identification and characterization of 22 genes that affect the vulval cell lineages of *Caenorhabditis elegans*. *Genetics* 110:17–72.
- Ferguson, E. L., P. W. Sternberg, and H. R. Horvitz. 1987. A genetic pathway for the specification of the vulval cell lineages of *Caenorhabditis elegans*. *Nature (London)* 326:259–267.
- Fields, C. 1990. Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. *Nucleic Acids Res.* 18:1509–1512.
- Fouser, L. A., and J. D. Friesen. 1987. Effects on mRNA splicing of mutations in the 3' region of the *Saccharomyces cerevisiae* actin intron. *Mol. Cell. Biol.* 7:225–230.
- Gao, F., H. Endo, and M. Yamamoto. 1989. Length heterogeneity in rat salivary gland $\alpha 2\mu$ globulin mRNAs: multiple splice-acceptors and polyadenylation sites. *Nucleic Acids Res.* 17:4629–4636.
- Gilliland, G., S. Perrin, K. Blanchard, and H. F. Bunn. 1990. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* 87:2725–2729.
- Gilliland, G., S. Perrin, and H. F. Bunn. 1990. Competitive PCR for quantitation of mRNA, p. 60–69. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), *PCR protocols: a guide to methods and applications*. Academic Press, Inc., San Diego, Calif.
- Goguel, V., X. Liao, B. C. Rymond, and M. Rosbash. 1991. U1 snRNP can influence 3'-splice site selection as well as 5'-splice site selection. *Genes Dev.* 5:1430–1438.
- Guillermi, H., P. Fanen, and C. Ferec. 1990. A 3' splice site consensus sequence mutation in the cystic fibrosis gene. *Hum. Genet.* 85:450–453.
- Hanioka, N., S. Kimura, U. A. Meyer, and F. J. Gonzalez. 1990. The human CYP2D locus associated with a common genetic defect in drug oxidation: a G₁₉₃₄ → A base change in intron 3 of a mutant CYP2D6 allele results in an aberrant 3' splice recognition site. *Am. J. Hum. Genet.* 47:994–1001.
- Hata, A., M. Emi, G. Luc, A. Basdevant, P. Gambert, P.-H. Iverius, and J.-M. Lalouel. 1990. Compound heterozygote for lipoprotein lipase deficiency: ser → thr²⁴⁴ and transition in 3' splice site of intron 2 (AG → AA) in the lipoprotein lipase gene. *Am. J. Hum. Genet.* 47:721–726.
- Hawkins, J. D. 1988. A survey of intron and exon lengths. *Nucleic Acids Res.* 21:9893–9907.
- Hedley, M. L., J. Forman, and P. W. Tucker. 1989. Mutation of 3' splice sites in two different class I genes results in different usage of cryptic splice sites. *J. Immunol.* 143:1018–1025.
- Higashi, Y., A. Tanae, H. Inoue, T. Hiromasa, and Y. Fujikuriyama. 1988. Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: possible gene conversion products. *Proc. Natl. Acad. Sci. USA* 85:7486–7490.
- Jackson, I. J. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* 19:3795–3798.
- Jongeward, G., R. V. Aroian, and P. W. Sternberg. Unpublished results.
- Jongeward, G., and P. W. Sternberg. Unpublished results.
- Kawasaki, E. S. 1990. Amplification of RNA, p. 21–27. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), *PCR protocols: a guide to methods and applications*. Academic Press, Inc., San Diego, Calif.
- Kay, R. J., R. H. Russnak, D. Jones, C. Mathias, and E. P. M. Candido. 1987. Expression of intron-containing *C. elegans* heat shock genes in mouse cells demonstrates divergence of 3' splice site recognition sequences between nematodes and vertebrates, and an inhibitory effect of heat shock on the mammalian splicing apparatus. *Nucleic Acids Res.* 15:3723–3741.
- Kobayashi, K., M. J. Jackson, D. B. Tick, W. E. O'Brien, and A. L. Beaudet. 1990. Heterogeneity of mutations in arginosuccinate synthetase causing human citrullinemia. *J. Biol. Chem.* 265:11361–11367.
- Kramer, J. M., G. N. Cox, and D. Hirsh. 1982. Comparisons of

- the complete sequences of two collagen genes from *Caenorhabditis elegans*. *Cell* 30:599-606.
32. Kramer, J. M., R. P. French, E. Park, and J. J. Johnson. 1990. The *Caenorhabditis elegans* *rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. *Mol. Cell. Biol.* 10:2081-2089.
 33. Kramer, J. M., J. J. Johnson, R. S. Edgar, C. Basch, and S. Roberts. 1988. The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* 55:555-565.
 34. Kuiper, M. T., M. Holtrop, H. Vennema, A. M. Lambowitz, and H. de Vries. 1988. A 3' splice site mutation in a nuclear gene encoding a mitochondrial ribosomal protein in *Neurospora crassa*. *J. Biol. Chem.* 263:2848-2852.
 35. Lamond, A. I., M. M. Konarska, and P. A. Sharp. 1987. A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.* 1:532-543.
 - 35a. Levy, A., and J. Kramer. Unpublished data.
 36. Metherall, J. E., F. S. Collins, J. Pan, S. M. Weissman, and B. G. Forget. 1986. β^0 Thalassemia caused by a base substitution that creates an alternative splice acceptor site in an intron. *EMBO J.* 5:2551-2557.
 37. Miner, J. H., and B. J. Wold. 1991. *c-myc* inhibition of MyoD and myogenin-initiated myogenic differentiation. *Mol. Cell. Biol.* 11:2842-2851.
 38. Mitchell, P. J., G. Urlaub, and L. Chasin. 1986. Spontaneous splicing mutations at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.* 6:1926-1935.
 39. Mules, E. H., C. E. Dowling, M. B. Petersen, H. H. Kazazian, and G. H. Thomas. 1991. A novel mutation in the invariant AG of the acceptor splice site of intron 4 of the β -hexosaminidase α -subunit gene in two unrelated American Black G_{M2} -Gangliosidosis (Tay-Sachs disease) patients. *Am. J. Hum. Genet.* 48:1181-1185.
 40. Nelson, K. K., and M. R. Green. 1988. Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes Dev.* 2:319-329.
 41. Newman, A. J., and C. Norman. 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* 68:743-754.
 42. Ogg, S. C., P. Anderson, and M. P. Wickens. 1990. Splicing of a *C. elegans* myosin pre-mRNA in a human nuclear extract. *Nucleic Acids Res.* 18:143-149.
 43. Patterson, B., and C. Guthrie. 1991. A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell* 64:181-187.
 44. Reed, R. 1989. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* 3:2113-2123.
 45. Reed, R., and T. Maniatis. 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* 41:95-105.
 46. Robberson, B. L., G. J. Cote, and S. M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10:84-94.
 47. Ruffner, D. E., and A. Dugaiczky. 1988. Splicing mutation in human hereditary analbuminemia. *Proc. Natl. Acad. Sci. USA* 85:2125-2129.
 48. Ruskin, B., and M. R. Green. 1985. Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. *Nature* (London) 317:732-734.
 49. Satokata, I., K. Tanaka, N. Miura, I. Miyamoto, Y. Satoh, S. Kondo, and Y. Okada. 1990. Characterization of a splicing mutation in group A xeroderma pigmentosum. *Proc. Natl. Acad. Sci. USA* 87:9908-9912.
 50. Senapathy, P., M. B. Shapiro, and N. L. Harris. 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* 183:252-278.
 51. Smith, C. W. J., E. B. Porro, J. G. Patton, and B. Nadal-Ginard. 1989. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* (London) 342:243-247.
 52. Su, T.-S., and L.-H. Lin. 1990. Analysis of a splice acceptor site mutation which produces multiple splicing abnormalities in the human arginosuccinate synthetase locus. *J. Biol. Chem.* 265:19716-19720.
 53. Sulston, J., and J. Hodgkin. 1988. Methods, p. 587-606. In W. B. Wood (ed.), *The nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 54. Sulston, J. E., and S. Brenner. 1974. The DNA of *Caenorhabditis elegans*. *Genetics* 77:95-104.
 55. Thomas, J., K. Lea, E. Zucker-Aprison, and T. Blumenthal. 1990. The spliceosomal snRNAs of *Caenorhabditis elegans*. *Nucleic Acids Res.* 18:2633-2642.
 56. Thomas, J. A., R. C. Conrad, and T. Blumenthal. 1988. The *C. elegans* trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. *Cell* 54:533-539.
 57. Treisman, R., S. H. Orkin, and T. Maniatis. 1983. Specific transcription and RNA splicing defects in five cloned β -thalassaemia genes. *Nature* (London) 302:591-596.
 58. Tromp, G., and D. J. Prockop. 1988. Single base mutation in the pro α 2(I) collagen gene that causes efficient splicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame pro α 2(I) chain. *Proc. Natl. Acad. Sci. USA* 85:5254-5258.
 59. Vijayraghavan, U., R. Parker, J. Tamm, Y. Iimura, J. Rossi, J. Abelson, and C. Guthrie. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *EMBO J.* 5:1683-1695.
 60. von Mende, N., D. Bird, P. S. Albert, and D. L. Riddle. 1988. *dpy-13*: a nematode collagen gene that affects body shape. *Cell* 55:567-576.
 61. Wenger, R. H., and P. J. Nielsen. 1991. Reannealing of artificial heteroduplexes generated during PCR-mediated genetic isotyping. *Trends Genet.* 7:178.
 62. Wiebauer, K., J.-J. Herrero, and W. Filipowicz. 1988. Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals. *Mol. Cell. Biol.* 8:2042-2051.
 63. Wieringa, B., E. Hofer, and C. Weissmann. 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β -globin intron. *Cell* 37:915-925.
 64. Zorn, A. M., and P. A. Krieg. 1991. PCR analysis of alternative splicing pathways: identification of artifacts generated by heteroduplex formation. *BioTechniques* 11:180-184.